

# Sense codon emancipation for proteome-wide incorporation of noncanonical amino acids: rare isoleucine codon AUA as a target for genetic code expansion

Nina Bohlke & Nediljko Budisa

Department of Chemistry, TU Berlin, Berlin, Germany

**Correspondence:** Nediljko Budisa,  
Department of Chemistry, TU Berlin,  
Biocatalysis Group, Mueller-Breslau-Straße  
10, D-10623 Berlin, Germany. Tel.: +49 30  
314 28 821; fax: + 49 30 314 28 279;  
e-mail: budisa@biocat.tu-berlin.de

Received 29 October 2013; revised 20  
December 2013; accepted 20 December  
2013. Final version published online 27  
January 2014.

DOI: 10.1111/1574-6968.12371

Editor: Jeff Cole

## Keywords

genetic code; sense codons; reassignment;  
orthogonal pairs; synthetic biology; codon  
emancipation.

## Abstract

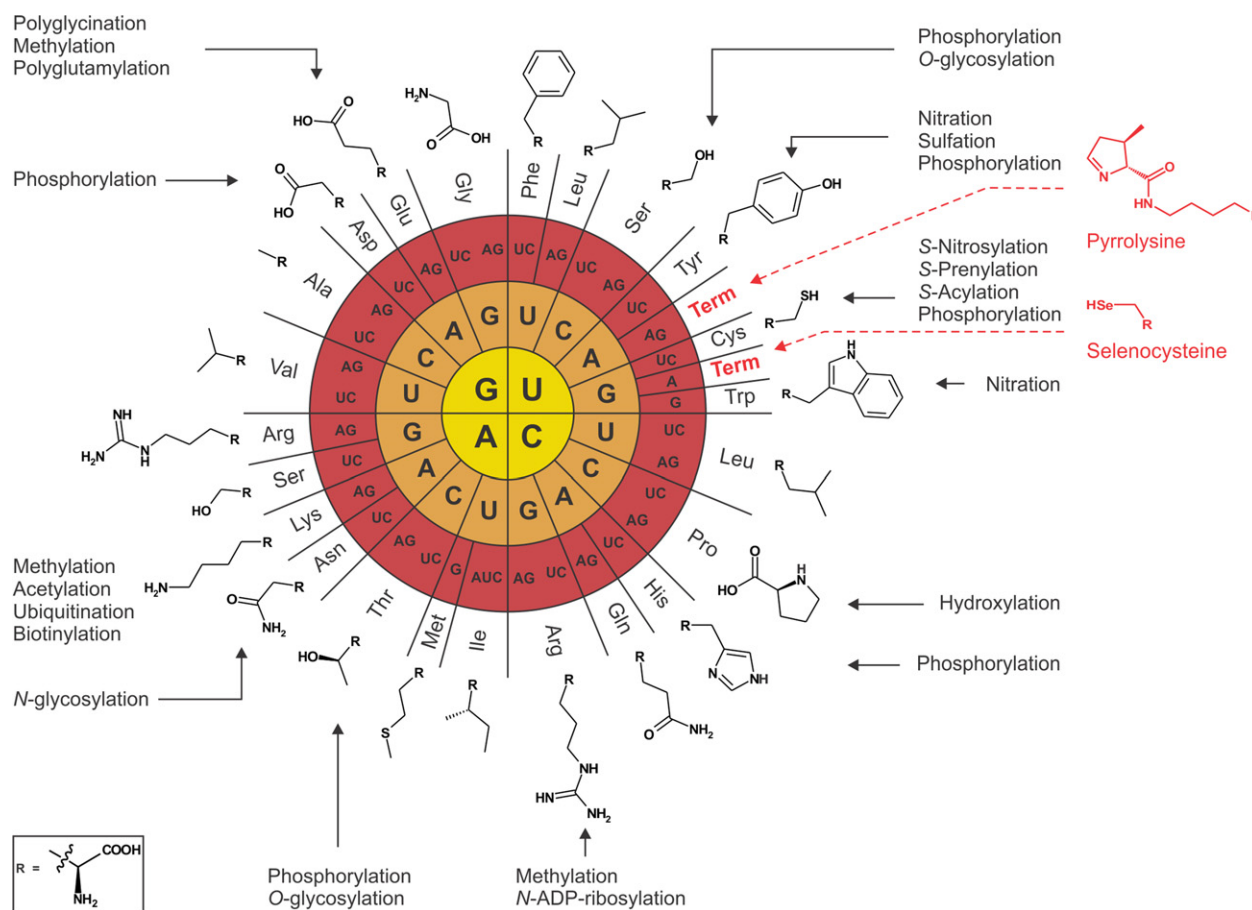
One of the major challenges in contemporary synthetic biology is to find a route to engineer synthetic organisms with altered chemical constitution. In terms of core reaction types, nature uses an astonishingly limited repertoire of chemistries when compared with the exceptionally rich and diverse methods of organic chemistry. In this context, the most promising route to change and expand the fundamental chemistry of life is the inclusion of amino acid building blocks beyond the canonical 20 (i.e. expanding the genetic code). This strategy would allow the transfer of numerous chemical functionalities and reactions from the synthetic laboratory into the cellular environment. Due to limitations in terms of both efficiency and practical applicability, state-of-the-art nonsense suppression- or frameshift suppression-based methods are less suitable for such engineering. Consequently, we set out to achieve this goal by sense codon emancipation, that is, liberation from its natural decoding function – a prerequisite for the reassignment of degenerate sense codons to a new 21st amino acid. We have achieved this by redesigning of several features of the post-transcriptional modification machinery which are directly involved in the decoding process. In particular, we report first steps towards the reassignment of 5797 AUA isoleucine codons in *Escherichia coli* using efficient tools for tRNA nucleotide modification pathway engineering.

## The genetic code: degeneracy, wobbling and codon usage

The nearly identical assignment of triplet codons to amino acids in all living organisms illustrates the universality of the genetic code and affirms the existence of a common origin. With only very few exceptions, ribosomal synthesis is restricted to 20 naturally occurring amino acids. Codon assignment in living cells relies on highly selective aminoacyl-tRNA synthetases (aaRS) which recognize both the amino acid to be charged and the related tRNA. By catalyzing a specific aminoacylation reaction, these enzymes provide the critical link between amino acids and codons and thereby establish the genetic code. As a general rule in living cells, one aminoacyl-tRNA synthetase exists for each canonical amino acid.

The most remarkable feature of the genetic code is its degeneracy. Of 64 possible triplet combinations, 61 sense

codons encode the 20 amino acids, while three stop codons (UAA, ochre; UGA, opal; and UAG, amber) are used to terminate translation (see Fig. 1). With the exceptions Met and Trp, most amino acids are thus represented by more than one codon. On average, each amino acid is encoded by three different codons. For example, Ile is encoded by AUC, AUU, and AUA. Surprisingly, these three codons are decoded by only two different tRNA molecules. The reason for this is wobble base pairing of the third codon position as proposed by Crick (1966). Different triplets encoding the same amino acid often vary only in the third position as evident from the aforementioned Ile codons. In this case, the first two codons, AUC and AUU, are decoded by the same anticodon GAU, either involving a G-C Watson-Crick base pairing or a G-U wobble base pairing. In general, the first two bases of a triplet form Watson-Crick base pairings, whereas the third position is able to establish wobble base pairings. In case



**Fig. 1.** Presentation of the genetic code in RNA format. Sixty-one triplets are assigned to 20 canonical amino acids. The three triplets UGA, UAA, and UAG are used as termination codons. All amino acids (except Met and Trp) are represented by more than one codon. The majority of proteins and peptides reach their active conformation after being post-translationally modified (PTM) or processed. Only a small portion of cellular proteins is changed in a co-translational manner, involving special proteinogenic amino acids such as selenocysteine and pyrrolysine. For example, most of all known regulatory and localization signals (e.g. phosphorylation and glycosylation) are generated by post-translational covalent side-chain modifications. These processes, catalyzed by enzymes and enzyme assemblies, are strictly separated from translation and represent events highly coordinated in time and space. For instance, in contrast to serine, its noncanonical counterpart phosphoserine, although frequently found in proteins, does not itself participate in translation (Adapted from Michael Hösl).

of the third Ile codon AUA, the first base (wobble position) of the anticodon CAT is post-transcriptionally modified with a lysine to form lysidine (L), which enables A–L wobble base pairings (details are described below).

Another important feature of the genetic code is the codon usage, the abundance of a codon in a particular DNA sequence, which can vary significantly between different organisms. For Ile decoding by *Escherichia coli*, AUU and AUC are frequently used, while the AUA codon is rarely used. A comparison of AUA codons used per 1000 bases in *E. coli* (3.7), *Saccharomyces cerevisiae* (17.8), and *Homo sapiens* (7.5) highlights significant differences (Nakamura *et al.*, 2000).

At the time of its deciphering, the genetic code was suggested to be universal in all organisms. Crick proposed a first theory on its evolution saying that the genetic code

was a result of a ‘frozen accident’ unable to evolve further, since ‘no new amino acid could be introduced without disrupting too many proteins’ (Crick, 1968; Söll & RajBhandary, 2006).

## Natural variations of the genetic code

The theory of a frozen universal genetic code is believed to be disproved by genetic code variations found in vertebrate mitochondria, where AUA encodes Met instead of Ile and UGA encodes Trp instead of being a translational termination signal (Barrell *et al.*, 1979; Osawa *et al.*, 1992). Today, around 20 variations of the standard genetic code are known, not only in organelles but also in free-living microorganisms (Ambrogelly *et al.*, 2007; Campbell *et al.*, 2013; O’Donoghue *et al.*, 2013).

Well-documented natural variations suggest the feasibility of genetic code expansion. The design of organisms capable to decode more than the common 20 building blocks could lead to proteins or even organisms with novel activities (Budisa, 2005). One of the earliest discovered variations of the genetic code were found to be deviations in codon assignment between nuclear and mitochondrial genes (Barrell *et al.*, 1979; Himeno *et al.*, 1987; Osawa *et al.*, 1992; Wolstenholme, 1992; Söll & RajBhandary, 1995).

Deviations in the codon meanings, especially in various mitochondrial genomes, clearly indicate that in the context of the smaller ('minimal') genome and proteome, such reassignments in the code structure are more likely, as lethal effects are less possible (Budisa, 2005). In other living beings, species-specific reassignments are sometimes employed as beneficial adaptive variations in a particular environment. One example is the Leu→Ser change for the CUG codon in some *Candida* species which probably leads to better survival of these mainly parasitic species in their hosts (Tuite & Santos, 1996). Disturbing effects of such changes are avoided in these cells as CUG is usually not used as a codon in the cellular mRNA. This reassignment is 'reserved' for specific genomic islands of this *Candida* pathogen; 'island genes' are usually involved in stress response (like heat shock) and pathogenicity (Suzuki *et al.*, 1997; Santos *et al.*, 2004).

Changes in codon meaning can be classified into two general categories. The first category includes reassignment of standard termination codons (UGA, UAA, and UAG) to Trp or Gln in some prokaryotes, archaea, mitochondria of several organisms, and even in nuclear genes of some protozoa (e.g. in *Mycoplasma mobile* the UGA termination codon encodes Trp). The second category of codon reassignment includes altered meanings in mitochondrial sense codons such as Ile→Met (Barrell *et al.*, 1979; Budisa, 2005). The mitochondrial AUA codon is deciphered by the tRNA<sup>Met</sup> containing an unusual modified nucleotide 5-formylcytidine (f<sup>5</sup>C) in the first position (number 34) of their anticodon. This tRNA translates both AUG and AUA as methionine (Takemoto *et al.*, 2009). Sixty-two such reassignments are known in mitochondria, but only one in nuclear genes: CUG Leu→Ser in *Candida cylindrical* (see above) (Osawa *et al.*, 1992; Tuite & Santos, 1996; Suzuki *et al.*, 1997; Santos *et al.*, 2004). As they do not affect the amino acid repertoire, such reassignments might be seen as flexibility in the 'frozen' structure of the genetic code.

To date, two examples for a natural expanded genetic code are known. An in-frame UGA termination codon also encodes selenocysteine (Sec), the 21st cotranslationally inserted amino acid. This recoding mechanism requires a tRNA<sub>UCA</sub>, a specialized translation elongation factor

(SelB) and a particular mRNA stem-loop structure, known as the seleocysteine insertion element (SECIS) (Söll, 1988; Schön *et al.*, 1989; Cusack *et al.*, 2005; Ambrogelly *et al.*, 2007). UAG is ambiguous in *Methanosarcinaceae* where, in addition to serving as a termination codon, it also encodes pyrrolysine (Pyl), the 22nd cotranslationally inserted amino acid. In this case, a new tRNA synthetase, pyrrolysyl-tRNA synthetase (PylRS), is essential for this recoding event (Söll, 1988; Hao *et al.*, 2002; Srinivasan *et al.*, 2002; Blight *et al.*, 2004; Polycarpo *et al.*, 2004).

In fact, no known examples for codon reassignment introduce a noncanonical amino acid in response to a sense codon.

## Experimental variations of the genetic code

As part of a synthetic genetic code, amino acids can be replaced by noncanonical (ncAA) ones (engineering), or alternatively, an ncAA can be added to the natural repertoire (expansion). Genetic code engineering commonly involves the use of wild-type aminoacyl-tRNA synthetases (aaRSs) to incorporate ncAAs that are close structural analogs of canonical amino acids. In this approach, a strain auxotrophic for one canonical amino acid (cAA) is used for substitution of the cAA with a structurally similar ncAA analog. In the newly synthesized proteins, the canonical amino acid is efficiently replaced by its analog at all sites (global replacement). This method is often called residue-specific incorporation, as all codons coding for a given amino acids are reassigned (Budisa, 2004). Basically, genetic code engineering leads to proteins containing 19 cAA and one ncAA or 18 cAA and 2 ncAA. Up to three different ncAA incorporated into a single protein were reported with this method (Lepthien *et al.*, 2010). Wang *et al.* (2004) showed that one mRNA can be translated in different ways depending on the relative rates of competing aminoacylation reactions. Depending on the intracellular ncAA levels and whether the host is engineered with an enhanced ValRS or IleRS activity, (2S, 3R)-4, 4-, 4-Trifluorovaline could be assigned either to isoleucine or valine codons (Wang *et al.*, 2004). Using this genetic engineering approach, proteins with new activities and properties can be generated, but the amino acid repertoire is still restricted to 20 amino acids.

Genetic code expansion allows for the incorporation of a 21st amino acid. This is usually achieved by assignment of a 'blank' codon (nonsense or frameshift suppression). This method enables the site-specific incorporation of an ncAA within the growing polypeptide chain, with the mRNA containing a nonsense codon at the desired position. This is realized by the use of an orthogonal suppressor aaRS:tRNA pair (o-pair) capable to decode

the nonsense codon. To date, various orthogonal aaRS: tRNA pairs have been developed (Liu & Schultz, 2010) with the TyrRS from *Methanocaldococcus jannaschii* (*mj*TyrRS) being most frequently used (Wang *et al.*, 2001). More recently, pyrrolysine aaRS (PylRS) and its cognate suppressor tRNA *pylT* from *Methanosarcinaceae* species have been shown to be highly potent orthogonal ncAA systems (Mukai *et al.*, 2008; Neumann *et al.*, 2008; Huang *et al.*, 2010). In contrast to *mj*TyrRS, the 'natural PylRS:tRNA o-pair' already displays a broad substrate tolerance towards orthogonal ncAAs, but PylRS does not utilize the common 20 amino acids as substrates (Polycarpo *et al.*, 2006; Li *et al.*, 2009). Within a short time, a large number of different PylRS variants for incorporation of versatile ncAAs were developed (Fekner *et al.*, 2010).

Nonsense suppression and frameshift suppression are widely used in academic research, although these approaches are still limited by several means. Both typically result in low protein yields due to cellular toxicity of the ncAA, suffer from context effects, and compete with the highly specialized endogenous termination machinery (Doerig *et al.*, 1988; Lepthien *et al.*, 2010; Young *et al.*, 2010). A detailed description about genetic code engineering and expansion is given elsewhere (Wang & Schultz, 2004; Wang *et al.*, 2006; Liu & Schultz, 2010; Antonczak *et al.*, 2011; Hoesl & Budisa, 2012).

## Sense codon-specific incorporation of ncAAs

Ultimately, the incorporation of an ncAA typically involves outcompeting the natural aaRS substrate, recoding a stop codon, or the use of four to five base codons. As most amino acids are decoded by more than one codon (exceptions: Trp and Met), our future aim is to emancipate a codon from its natural meaning and to reassign it to an ncAA. This would allow proteome-wide incorporation of a 21st amino acid in response to a sense codon.

It is assumed that between 30 and 40 sense codons are required to encode the genetic information of an organism. Therefore, a large number of sense codons (> 20) should be available for recoding experiments (Krishnakumar *et al.*, 2013). Correspondingly, the genetic code would be substantially expanded by such proteome-wide replacements. To our understanding, a synthetic organism with an expanded genetic code is not able to live without the ncAA, as the newly reassigned sense codon becomes not assigned.

One step towards sense codon reassignment was described by Kwon *et al.* (2003). In *E. coli*, Phe is naturally encoded by UUC and UUU codons, and a single tRNA<sup>Phe</sup><sub>GAA</sub> decodes both via Watson–Crick base pairing and wobble base pairing, respectively. Introduction of a heterologous

tRNA with an AAA anticodon (ytRNA<sup>Phe</sup><sub>AAA</sub>) and a corresponding aaRS (*yaaRS*) from *S. cerevisiae* resulted in competition between the two tRNAs for the UUU codon. The synthetic site of *yPheRS* was engineered to activate the ncAA 2-naphthylalanine. Depending on experimental setup, assignment of UUC to Phe was quantitative and UUU translation into 2-naphthylalanine reached 80% (Kwon *et al.*, 2003). Nevertheless, the cells were still able to incorporate Phe at UUU positions. In the absence of the ncAA, the cells could easily grow using their endogenous tRNA<sup>Phe</sup><sub>GAA</sub> for the incorporation of Phe at UUU positions. A complete sense codon-specific reassignment would lead to cell death in the absence of the ncAA. A nice overview about reassignment of sense codons *in vivo* and corresponding methods is given by Link & Tirrell (2005) and extensively discussed by Budisa (2005, pp. 97–100).

## Choice of the degenerate codon: AUA as a target for reassignment

Real expansion of the genetic code will take place once when we would be able to add synthetic amino acids in a proteome-wide manner by genome-wide reassignments of target codons. Here, we elaborate the concept and report first experimental steps towards the emancipation/liberation of degenerate sense codons from their canonical function, a step towards estranging the genetic code that should enable addition of novel amino acids to the existing repertoire.

For the choice of the sense codon to be reassigned, two aspects were considered highly important. The first important point is the presence of a specific modification pathway, which is necessary for the decoding process of the codon. Through the selective interruption of this modification, the decoding mechanism can be eliminated. Thus, the codon is liberated from its original meaning and is available as an open sense codon. Now, the certain codon can be reassigned to a noncanonical amino acid. To minimize disrupting effects through the appearance of an ncAA in the whole proteome, it is favorably to select a codon rarely used. The second important selection criterion is the choice of a noncanonical amino acid whose proteome-wide replacement will be tolerated by the cell. Therefore, choosing a codon which encodes an amino acid with little metabolic involvement and catalytic functionality might be a good starting point. In the genetic code, all coding triplets with a central U (the so-called XUX group of codons) are cognate to apolar (hydrophobic) amino acids, for example Leu, Ile, and Val. These residues are chemically inert and rarely directly involved in the catalytic function of proteins, but frequently play an essential role in protein folding as well as in binding/recognition of hydrophobic ligands. Their synthetic



chemical analogs such as trifluoroleucine have been known for more than 50 years to be translationally competent, allowing their incorporation into the nascent polypeptide chain (Rennert & Anker, 1963).

With Ile rarely involved in catalytic reactions, the rare codon AUA perfectly fits for reassignment experiments. In *E. coli*, the codon usage of AUA is low. With only 5797 occurrences in the genome, it is usually involved in regulatory events such as ribosomal pausing (Lajoie *et al.*, 2013a). Consequently, we speculated that replacement of Ile by ncAA at AUA positions might be tolerated by the *E. coli* cell.

Next, codon ambiguity has widespread effects on protein function mainly resulting in detrimental effects on cellular fitness. However, these growth defects can be compensated by multiple mutations in various proteins by improving their ability to grow in the presence of the ncAA. For example, Philippe Marliere and coworkers generated *E. coli* mutant strains capable of installing translational pathways specific for the incorporation of additional amino acids into proteins *in vivo* (Lemeignan *et al.*, 1993). Their selection procedure resulted in stable bacterial strains with coding triplets that can be read ambiguously. Similarly, Döring & Marliere (1998) constructed stable bacterial strains with ambiguous reading of the rare coding triplet AUA by overexpression of CysRS:tRNA<sup>Cys</sup>. Experimentally designed selective pressure yielded almost full reassignment (Ile→Cys) of the AUA codon in *E. coli*. Such proteome-wide substitutions were possible as Ile side chains evidently do not act as catalytic residues in any enzyme of the *E. coli* proteome (Döring & Marliere, 1998; Budisa, 2005). Although the repertoire of amino acids was not enlarged, it could be shown that Cys miscoded at the Ile AUA codon was well tolerated by the cell and that codon reading as part of normal translation can be altered *in vivo*. Döring *et al.* (2001) were also able to globally substitute Val by the ncAA aminobutyric acid (Abu) in a quarter of the cellular proteome using an *E. coli* strain with defective editing function of ValRS (Döring *et al.*, 2001). Furthermore, a nonreverting strain of *E. coli* with impaired editing function has been generated. Under Ile starvation conditions and relative to a wild-type *E. coli* strain, it achieves higher culture growth in combination with analogs such as norvaline (Pezo *et al.*, 2004). Later, Bacher *et al.* (2005) abolished the editing function of IleRS and found that bacteria grow slowly even with half of their isoleucine residues being miscoded. Taken together, the rare Ile AUA codon seems to be a feasible target for sense codon reassignment experiments.

Finally, proteome-wide addition of noncanonical amino acids associated with breaking the degeneracy of the genetic code should be tolerated by correspondingly engineered cells, in an ideal case by setting up long-term

cultivation experiments (Lenski *et al.*, 1991). In the Lenski experiments or using automated devices (de Crécy-Lagard *et al.*, 2001), such genetic estrangement through codon reassignment associated with the use of ncAAs as building blocks would be enforced towards propagation and selection of robust and viable cells with built-in alien nutrients.

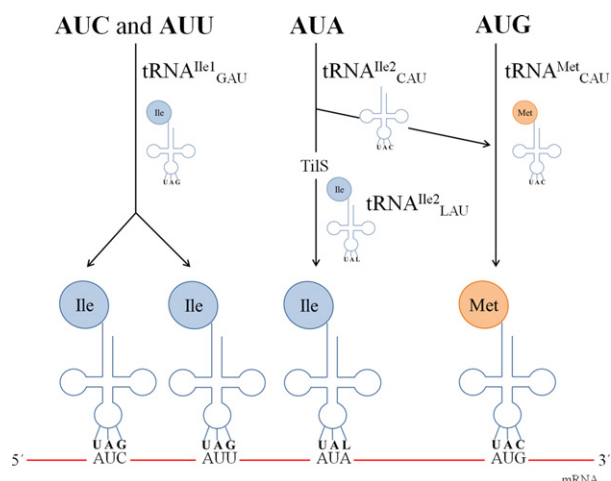
## Rare isoleucine AUA codon decoding in *E. coli*

With great accuracy, tRNA isoacceptors are aminoacylated with their related amino acid *in vivo*. In *E. coli*, 49 tRNAs function to read the sense codons of the messenger RNA, converting them into 20 amino acids of the protein sequences (Budisa, 2004). Tuning of tRNA-mediated translation has been enabled by more than 70 chemically distinct post-transcriptional base modifications in tRNA species (Agris *et al.*, 2007).

In *E. coli*, isoleucine is decoded by three different codons. The two frequently used codons, AUC and AUU, are translated by the high-abundance tRNA species, tRNA<sup>Ile1</sup>, which harbors a G34AU anticodon and pairs with the two codons. The rare AUA codon is translated by low-abundance tRNA<sup>Ile2</sup> that carries the same anticodon (CAU) as tRNA<sup>Met</sup>. Together, the tRNA<sup>Ile1</sup> and tRNA<sup>Ile2</sup> pair is able to decipher all Ile codons accurately (Muramatsu *et al.*, 1988a, b; Soma *et al.*, 2003; Ikeuchi *et al.*, 2005; Nakanishi *et al.*, 2009).

Bacteria decode AUA via a tRNA in which the wobble position C34 is modified with a lysine, resulting in lysidine (L, 2-lysyl-cytidine). The enzyme catalyzing this modification is called lysidine-tRNA synthase (TilS) (Muramatsu *et al.*, 1988a; Soma *et al.*, 2003; Silva *et al.*, 2006). This lysine-containing cytidine derivative occurs in almost all bacteria (Harada & Nishimura, 1974; Muramatsu *et al.*, 1988b), whereas another chemical modification (2-agmatinylcytidine, agm2C) exists in archaea. The latter modification is introduced by agmatidine synthase TiaS (Ikeuchi *et al.*, 2010; Mandal *et al.*, 2010). The minor tRNA<sup>Ile2</sup> lacking the lysyl/agmatidine group on C34 is aminoacylated by MetRS and exclusively reads the Met codon AUG (Silva *et al.*, 2006). An overview about Ile decoding in *E. coli* is given in Fig. 2.

It was shown that EctRNA<sup>Ile2</sup> bearing an LAU anticodon is charged exclusively with Ile by IleRS and recognizes only the AUA codon. According to the chemical structure of lysidine (Fig. 3a), conjugation of the lysine moiety at the C2 position of cytidine induces a tautomeric conversion with protonation of the N3 position and imino group formation at the C4 position. By this modification, the proton donor-acceptor pattern of C in hydrogen bonding is completely altered towards the function of U,



**Fig. 2.** Ile decoding in *Escherichia coli*. Isoleucine is decoded by three different codons. The two frequently used codons AUC and AUU are translated by the high-abundance tRNA species, tRNA<sup>Ile1</sup>, which harbors a GAU anticodon. The rare AUA codon is translated by the low-abundance tRNA<sup>Ile2</sup> with the same anticodon CAU as tRNA<sup>Met</sup>. tRNA<sup>Ile2</sup> is recognized by MetRS when the anticodon at position C34 lacks a lysine modification catalyzed by an enzyme called TilS.

enabling L to pair with A but not with G (Suzuki & Miyauchi, 2010).

In summary, the L34 modification has a dual role: It causes the mature tRNA<sup>Ile2</sup> to be aminoacylated by IleRS and converts the codon specificity from AUG to AUA (Muramatsu *et al.*, 1988a, b) (Figure 3b).

## An essential enzyme: tRNA<sup>Ile</sup>-lysine synthetase

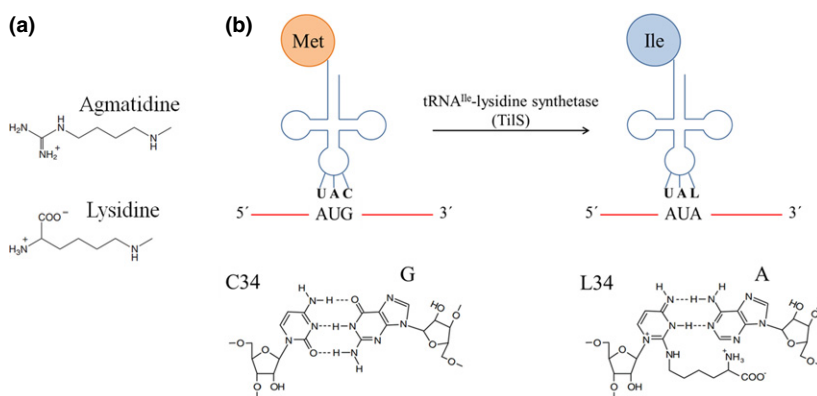
Soma *et al.* (2003) identified TilS as the enzyme responsible for the formation of lysidine. It could be shown that partial inactivation of *tilS* results in an AUA codon-dependent translational defect, which supports the notion that TilS is an RNA modifying enzyme that plays a critical role in the accurate decoding of the genetic code.

Complete TilS inactivation is lethal, as the AUA codon becomes an unassigned codon without its cognate decoding tRNA (Nakanishi *et al.*, 2009). Therefore, bacterial *tilS* was proposed to be one of the 206 essential protein-coding genes required for maintaining bacterial cell life (Gil *et al.*, 2004) and to be essential for reconstructing a minimal synthetic cell (Mushegian & Koonin, 1996; Moya *et al.*, 2009; Gibson *et al.*, 2010). Rescue of the translational defect caused by inactivation of TilS could be performed by a specialized tRNA able to decode the AUA codon. However, remember that codons within the Ile/Met decoding family box (e.g. AUA/AUG) are among the most difficult coding units to be discriminated accurately.

Nonetheless, the sequenced genomes of a few living organisms appear to lack a TilS homolog and tRNA<sup>Ile2</sup><sub>CAU</sub> (Jaffe *et al.*, 2004; Silva *et al.*, 2006), which suggest that alternative mechanisms of AUA translation exist. Bacterial genomes lacking *tilS* are found in *Bifido bacterium* (two species), *M. mobile*, and *Neorickettsia* (two species) (Fabret *et al.*, 2011). These species possess a U<sub>34</sub>-containing tRNA<sup>Ile</sup> (tRNA<sup>Ile3</sup><sub>UAU</sub>) for deciphering AUA codons. Notably, there are bacterial species that contain all three a U<sub>34</sub>-harboring tRNA<sup>Ile3</sup>, a *tilS*, and a C<sub>34</sub>-containing-tRNA<sup>Ile2</sup> gene pair in their genomes (*Lactobacillus*, *Rhodospirillum*, *Planctomyces*, *Ktedonobacter*, *Cyanothecae*) (Fabret *et al.*, 2011). In this way, these species can be regarded as evolutionary intermediates between the predominant group of TilS-containing organisms and the rare group of organisms lacking the functional pair of *tilS/tiaS* and tRNA<sup>Ile2</sup><sub>CAU</sub> genes. In other words, they contain a temporary ambiguous decoding system – a prediction of the ‘ambiguous intermediate’ theory (Schultz & Yarus, 1994a, b; Jones *et al.*, 2008; Fabret *et al.*, 2011).

## *Mycoplasma mobile* tRNA<sup>Ile3</sup><sub>UAU</sub>:IleRS pair is able to decipher AUA codons in *E. coli*

As mentioned before, *M. mobile* does not carry the *tilS* gene but a tRNA<sup>Ile3</sup><sub>UAU</sub> without any modification. The



**Fig. 3.** (a) Chemical structures of agmatidine and lysidine. (b) Lysidine converts both the codon and amino acid specificities of tRNA<sup>Ile</sup>. Modified from Suzuki & Miyauchi (2010).

acquisition of a tRNA<sup>Ile3</sup><sub>UAU</sub> gene by *M. mobile* has led to the loss of both the tRNA<sup>Ile2</sup><sub>CAT</sub> and the *tilS* gene (Silva *et al.*, 2006). Very recently, Taniguchi *et al.* (2013) demonstrated that *MmIleRS* recognizes the UAU anticodon without any modification, whereas *E. coli* IleRS did not efficiently aminoacylate tRNA<sup>Ile3</sup><sub>UAU</sub> *in vitro*. They found Arg865 in *MmIleRS* to be critical for the specificity to recognize the AUA codon and showed that when the corresponding site in *EcIleRS* W905 is mutated to Arg, *EcIleRS* was also able to charge tRNA with a UAU anticodon efficiently. Furthermore, they found that *MmtRNA*<sup>Ile3</sup><sub>UAU</sub> cannot distinguish between AUA and AUG codons *in vitro* on *E. coli* ribosomes. This finding suggests that in *E. coli*, tRNA<sup>Ile3</sup><sub>UAU</sub> might also read AUG codons as Ile to an extent that could lead to translational defects. As tRNA<sup>Met</sup><sub>CAU</sub> strongly recognizes the AUG codon in the cell, potential misreading might be reduced by competition of *MmtRNA*<sup>Ile3</sup><sub>UAU</sub> with endogenous tRNA<sup>Met</sup><sub>CAU</sub> (Taniguchi *et al.*, 2013).

### ***Escherichia coli* $\Delta$ *tilS* strain rescued with *MmtRNA*<sup>Ile3</sup><sub>UAU</sub>:IleRS capable to read rare AUA codons**

We have now tested the ability of *M. mobile* tRNA<sup>Ile3</sup><sub>UAU</sub>:IleRS pair (*MmPair*) to rescue the lethal *tilS* deletion in *E. coli* through direct AUA codon reading. The ability to decode the AUA codon with a heterologous tRNA<sup>Ile3</sup><sub>UAU</sub>:IleRS pair – which can later be modified to incorporate an ncAA at AUA positions – is the first step towards AUA sense codon reassignment.

To analyze whether the *MmtRNA*<sup>Ile3</sup><sub>UAU</sub>:IleRS pair (*MmPair*) is able to read the AUA codon in a *tilS*-depleted *E. coli* strain, we constructed several rescue vectors. First, tRNA<sup>Ile3</sup><sub>UAU</sub> and *MmIleRS* were cloned separately as well as together into plasmid pNB26'2. Furthermore, *tilS* was cloned into pNB26'1 (Figure S1) as a positive control. Both plasmids carry a p15A origin of replication and pLac promoter. These four different plasmids were used in *tilS* deletion experiments designed to restore AUA codon readout. The chromosomal *tilS* in *E. coli* MG1655 was deleted following the method of Datsenko & Wanner (2000). Obtained colonies were analyzed by PCR with different confirmation primers (Fig. 4a; Oligonucleotide sequences are given in Supporting Information, Table S1). For experimental details see Data S1.

For pNB26'1\_*tilS*, genomic *tilS* was successfully replaced by *cat* (chloramphenicol acetyl-transferase) as shown in Fig. 4b (1 and 2). Colonies obtained for pNB26'2\_*MmtRNA*<sup>Ile3</sup><sub>UAU</sub> and for pNB26'2\_*MmIleRS* retained the *tilS* gene and were most probably spontaneous Cm-resistant colonies (Figure 4b (2 and 3)). These findings suggest that *E. coli* IleRS is not able to charge

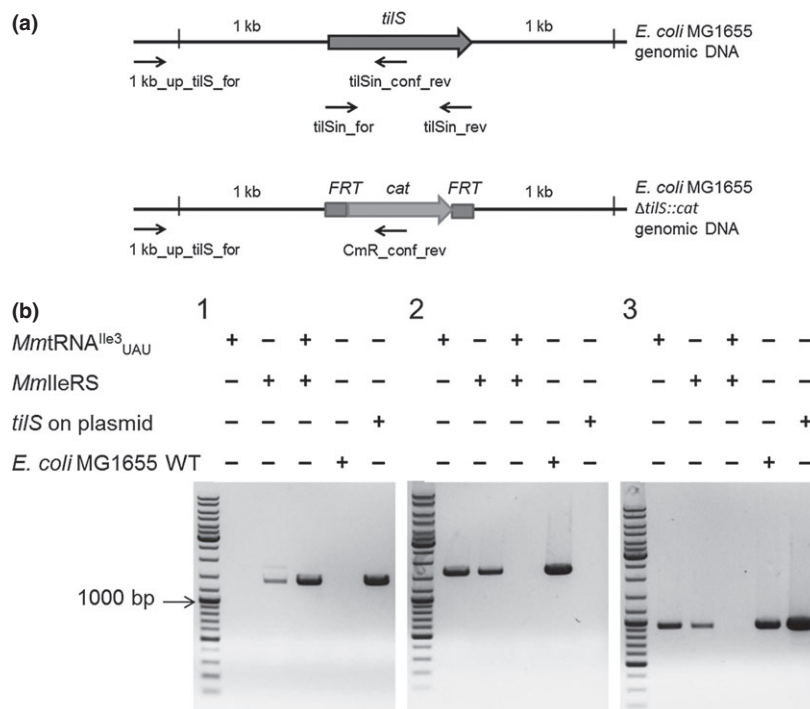
*MmtRNA*<sup>Ile3</sup><sub>UAU</sub> *in vivo*, which is consistent with the *in vitro* results of Taniguchi *et al.* (2013). Regarding the deletion experiment with pNB26'2\_*MmIleRS*, two explanations are plausible. First, *MmIleRS* might not be able to charge *EctRNA*<sup>Ile1</sup><sub>GAT</sub> and *EctRNA*<sup>Ile2</sup><sub>CAT</sub>. Alternatively, *MmIleRS* could charge one or both *EctRNAs*, whereas the aminoacylated tRNA is not able to decipher the AUA codon on the *E. coli* ribosome. As Taniguchi *et al.* (2013) could show *in vitro* aminoacylation of *EctRNA*<sup>Ile1</sup><sub>GAT</sub> by *MmIleRS*, the second option is most probably correct. Nevertheless, using tRNA or IleRS separately, it was not possible to completely remove *tilS* from the *E. coli* genome. In contrast, the combination of *MmtRNA*<sup>Ile3</sup><sub>UAU</sub> and *MmIleRS* was able to rescue the lethal *tilS* deletion in *E. coli* MG1655, rendering the cells viable without the essential enzyme *TilS*.

The generated *E. coli* MG1655  $\Delta$ *tilS*::*cat* strain (pNB26'2\_*MmPair*) grows slower than the wild-type equivalent (Fig. 5 and Fig. S2). The calculated doubling times are 38 and 28 min, respectively (Fig. S3). The lower growth rate might be due to interference with the AUG codon by *MmtRNA*<sup>Ile3</sup><sub>UAU</sub> or due to low intracellular levels of *MmtRNA*<sup>Ile3</sup><sub>UAU</sub>. Currently, we are investigating this phenomenon in more detail.

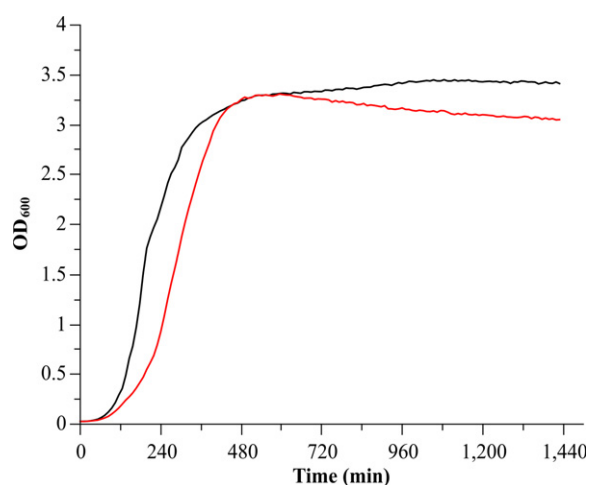
Taken together, our results show that it is possible to liberate the AUA codon from its natural deciphering mechanism in *E. coli*. Using *MmIleRS* and *MmtRNA*<sup>Ile3</sup><sub>UAU</sub>, we were able to restore AUA decoding and to rescue the lethal chromosomal *tilS* deletion (see Fig. 4). Figure 6 represents the suggested rescue mechanism of the *MmPair*. In future, we will engineer an orthogonal aaRS to specifically incorporate a 21st ncAA at AUA positions to generate an *E. coli* strain with a proteome-wide expanded genetic code through sense codon reassignment.

### **Outlook: towards genetic code expansion via experimental reassignment of AUA codons**

Currently available methodologies for genetic code expansion are well suited for small-scale experiments focused on particular academic questions (e.g. those using fluorescence tags) but are of quite limited use for the design and creation of robust artificial diversity in synthetic cells. Thus, we believe that it is necessary to replace the traditional suppression-based approaches with reassignments of sense (preferably rare) codons. However, we are well aware that rare codons such as AUA are often part of translational fine-tuning systems such as ribosomal pausing (Zhang *et al.*, 2009). Nonetheless, we succeeded to develop a rescue strategy that compensated for the translational defects by importing the AUA decoding system from *M. mobile*, which consists of *MmIleRS* and *MmtRNA*<sup>Ile3</sup><sub>UAU</sub>.



**Fig. 4.** (a) Confirmation primers used to verify *tilS* deletion in genomic DNA of *Escherichia coli* MG1655. Primer combination 1kb\_up\_tilS\_for and tilSin\_conf\_rev amplifies a part of *tilS* and 1 kb upstream of genomic *E. coli* DNA. Combination tilSin\_for and tilSin\_rev amplifies *tilS* in genomic or plasmid DNA. Combination 1kb\_up\_tilS\_for and CmR\_conf\_rev amplify part of *cat* and 1 kb upstream of genomic *E. coli* DNA to verify that *cat* replaced *tilS*. (b) PCR verification of  $\Delta tilS::cat$  genotype in *E. coli* MG1655. 1) PCR results with primers 1kb\_up\_tilS\_for and CmR\_conf\_rev. For single *MmtRNA*<sup>Ile3</sup><sub>UAU</sub>, no replacement of *tilS* with *cat* was observed. As a control, also *E. coli* MG1655 WT showed no PCR product. For *MmlleRS*, *tilS* on a plasmid and the combination of *MmtRNA*<sup>Ile3</sup><sub>UAU</sub> with *MmlleRS*, *tilS* replacement with *cat* is observed. 2) PCR results with primers 1kb\_up\_tilS\_for and tilSin\_conf\_rev. For all variants except *MmtRNA*<sup>Ile3</sup><sub>UAU</sub> together with *MmlleRS* and *tilS* on a plasmid, *tilS* was found not to be deleted. 3) PCR results with primers tilSin\_for and tilSin\_rev. These results show that only for the *MmtRNA*<sup>Ile3</sup><sub>UAU</sub>/*MmlleRS* combination, *E. coli* is able to live without *tilS*.

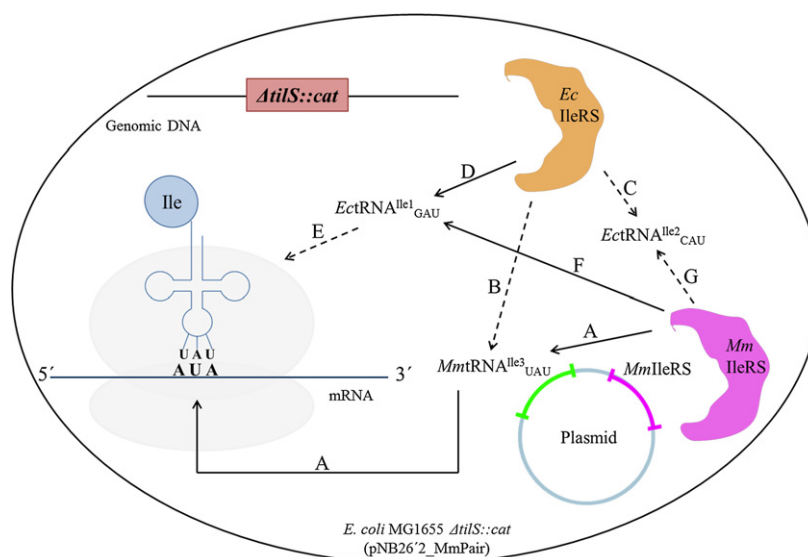


**Fig. 5.** Comparison of growth curves between *E. coli* MG1655 WT (black line) and *E. coli* MG1655  $\Delta tilS::cat$  (pNB26'2\_MmPair) (red line).

In the next step, we will use an engineered isoleucyl-tRNA synthetase (IleRS) from an evolutionary distant organism to reassign the AUA codon with an ncAA. Just like related enzymes for branched-chain amino acids (Thr, Val, Leu), this IleRS exhibits a hydrolytic editing that is capable to dissociate noncognate amino acids. In conjunction with the substrate recognition and activation by the active site architecture, this leads to the 'double sieve' concept (Fersht, 1977; Fersht & Dingwall, 1979). The availability of high-resolution X-ray crystal structures of different IleRSs (in apo-forms or in tRNA/ATP/amino acid complexes) is thus of enormous importance for the identification of residues crucial for amino acid recognition as well as for aaRS:tRNA interactions. The basic enzyme engineering strategy will rely on well-established mutagenesis approaches to redesign the binding pocket and to attenuate the aaRS editing activity. For example, Mursinna & Martinis (2002) reported a rational design that blocked the amino acid editing reaction in *EcLeuRS*. In particular, a



**Fig. 6.** AUA decoding in *Escherichia coli*  $\Delta$ *tilS::cat* (pNB26'2\_MmPair) cells. The MmPair (*Mmt*RS and *Mmt*RNA<sup>Ile3</sup><sub>UAU</sub>) is able to rescue the lethal chromosomal *tilS* deletion in *E. coli* (A). *Ect*RS is not able to charge *Mmt*RNA<sup>Ile3</sup><sub>UAU</sub> (B) and *Ect*RNA<sup>Ile2</sup><sub>CAU</sub> (C), whereas *Ect*RS is capable to charge *Ect*RNA<sup>Ile1</sup><sub>GAU</sub> (D), but this tRNA is not able to read the AUA codon at the ribosome (E). Finally, *Mmt*RS is able to charge *Ect*RNA<sup>Ile1</sup><sub>GAU</sub> (F), but not *Ect*RNA<sup>Ile2</sup><sub>CAU</sub> (G), as shown elsewhere (Taniguchi *et al.*, 2013).



single Thr at position 252 was found to play a critical role, with its mutation to bulky amino acids like Phe or Tyr resulting in the loss of the editing function (Mursinna & Martinis, 2002). Tang & Tirrell (2002) have identified three LeuRS mutants (Thr252Tyr, Thr252Leu, Thr252Phe) with attenuated editing activities that allow successful translation of norvaline, allylglycine, homoallylglycine, homopropargylglycine, and 2-butynylglycine at Leu positions of the model protein (Tang & Tirrell, 2002). This approach should be applicable to all 'double sieve' aaRS for branched-chain amino acids such as ThrRS, ValRS, LeuRS, and IleRS.

By a variety of mechanisms, long-term cultivation of bacteria can lead to increased mutation rates – the so-called adaptive mutation phenomenon (Marlière *et al.*, 2011). Under such conditions, these organisms need to explore both a larger genetic space and a large protein folding space to accommodate the desired ncAAs in their proteome and intermediary metabolism. Thus, we anticipate long-term cultivation experiments using a special turbidostat system as a promising tool for creating codon-emancipated cells (de Crécy-Lagard *et al.*, 2001). Via a controllable and defined selective pressure, codon removal will be actively enforced in microbial cells, allowing the reassignment of AUA codons to suitable Ile analogs. Subjected to successive selection cycles in continuous culture, *E. coli* strains with ambiguous reading of this codon will prevail. Eventually, sufficiently stable variants will be perpetuated.

We are fully aware that the reassignment of AUA codons to ncAAs potentially causes detrimental effects on cell survival due to their partial localization in essential genes. However, the same argument also applies to the elimination or reassignment of amber stop codons (UGA) in the genome of *E. coli* (321 ORFs). Nonetheless, the groups of

Church, Sakamoto, and Yokoyama have shown very recently that the latter codon can be either fully replaced via oligonucleotide-mediated MAGE technology or reassigned by simple knock out of release factor 1, (RF1), compensated by the presence of a suppressor tRNA capable to decode the UAG codon (Mukai *et al.*, 2010; Isaacs *et al.*, 2011; Johnson *et al.*, 2011). Together, the replacement of all UAG codons with UAA codons in *E. coli* and the deletion of RF1 allowed for the reassignment of UAG translation function (Lajoie *et al.*, 2013b). Recently, George Church and coworkers analyzed the possibility of replacing degenerate rare codons in essential genes with a synonymous codon. They selected 42 essential genes and removed all instances of 13 rare codons from these genes indicating that genome-wide removal of rare codons is feasible (Lajoie *et al.*, 2013a). In light of these results, our strategy of proteome-wide AUA codon reassignment appears feasible.

## Acknowledgements

We are indebted to Dr Tobias Baumann for critically reading this manuscript. We gratefully acknowledge financial support from the DFG Sonderforschungsbereich 765 (SFB 765; Multivalenz als chemisches Organisations- und Wirkprinzip), FP7 EU-funded METACODE-Consortium, UniCat Excellence Cluster of TU Berlin, and from DFG Forschergruppe 1805 (Ribosome Dynamics in Regulation of Speed and Accuracy of Translation).

## References

- Agris PF, Vendeix FA & Graham WD (2007) tRNA's wobble decoding of the genome: 40 years of modification. *J Mol Biol* **366**: 1–13.

- Ambrogelly A, Palioura S & Söll D (2007) Natural expansion of the genetic code. *Nat Chem Biol* **3**: 29–35.
- Antonczak AK, Morris J & Tippmann EM (2011) Advances in the mechanism and understanding of site-selective noncanonical amino acid incorporation. *Curr Opin Struct Biol* **21**: 481–487.
- Bacher JM, de Crécy-Lagard V & Schimmel PR (2005) Inhibited cell growth and protein functional changes from an editing-defective tRNA synthetase. *P Natl Acad Sci USA* **102**: 1697–1701.
- Barrell BG, Bankier AT & Drouin J (1979) A different genetic code in human mitochondria. *Nature* **282**: 189–194.
- Blight SK, Larue RC, Mahapatra A, Longstaff DG, Chang E, Zhao G, Kang PT, Green-Church KB, Chan MK & Krzycki JA (2004) Direct charging of tRNA(CUA) with pyrrolysine *in vitro* and *in vivo*. *Nature* **431**: 333–335.
- Budisa N (2004) Prolegomena to future experimental efforts on genetic code engineering by expanding its amino acid repertoire. *Angew Chem Int Ed* **43**: 6426–6463.
- Budisa N (2005) *Engineering the Genetic Code*. Wiley-VCH Verlag GmbH & Co, KGaA, Weinheim, Germany.
- Campbell JH, O'Donoghue P, Campbell AG, Schwientek P, Sczyrba A, Woyke T, Söll D & Podar M (2013) UGA is an additional glycine codon in uncultured SR1 bacteria from the human microbiota. *P Natl Acad Sci USA* **110**: 5540–5545.
- Crick FH (1966) Codon–anticodon pairing: the wobble hypothesis. *J Mol Biol* **19**: 548–555.
- Crick FH (1968) The origin of the genetic code. *J Mol Biol* **38**: 367–379.
- Cusack S, Francklyn C & Ibba M (2005) *The Aminoacyl-tRNA Synthetases*. Landes Bioscience, Georgetown, TX.
- Datsenko KA & Wanner BL (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *P Natl Acad Sci USA* **97**: 6640–6645.
- de Crécy-Lagard VA, Bellalou J, Mutzel R & Marlière P (2001) *BMC Biotechnol* **1**: 10.
- Doerig RE, Suter B, Gray M & Kubli E (1988) Identification of an amber nonsense mutation in the *rosy516* gene by germline transformation of an amber suppressor tRNA gene. *EMBO J* **7**: 2579–2584.
- Döring V & Marlière P (1998) Reassigning cysteine in the genetic code of *Escherichia coli*. *Genetics* **150**: 543–551.
- Döring V, Mootz HD, Nangle LA, Hendrickson TL, de Crécy-Lagard V, Schimmel P & Marlière P (2001) Enlarging the amino acid set of *Escherichia coli* by infiltration of the valine coding pathway. *Science* **292**: 501–504.
- Fabret C, Dervyn E, Dalmais B, Guillot A, Marck C, Grosjean H & Noirot P (2011) Life without the essential bacterial tRNA Ile2-lysine synthetase TilS: a case of tRNA gene recruitment in *Bacillus subtilis*. *Mol Microbiol* **80**: 1062–1074.
- Fekner T, Li X & Chan MK (2010) Pyrrolysine analogs for translational incorporation into proteins. *Eur J Org Chem* **2010**: 4171–4179.
- Fersht AR (1977) Editing mechanisms in protein synthesis. Rejection of valine by the isoleucyl-tRNA synthetase. *Biochemistry* **16**: 1025–1030.
- Fersht AR & Dingwall C (1979) Evidence for the double-sieve editing mechanism in protein synthesis. Steric exclusion of isoleucine by valyl-tRNA synthetases. *Biochemistry* **18**: 2627–2631.
- Gibson DG, Glass JI, Lartigue C *et al.* (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329**: 52–56.
- Gil R, Silva FJ, Peretó J & Moya A (2004) Determination of the core of a minimal bacterial gene set. *Microbiol Mol Biol Rev* **68**: 518–537, table of contents.
- Hao B, Gong W, Ferguson TK, James CM, Krzycki JA & Chan MK (2002) A new UAG-encoded residue in the structure of a methanogen methyltransferase. *Science* **296**: 1462–1466.
- Harada F & Nishimura S (1974) Purification and characterization of AUA specific isoleucine transfer ribonucleic acid from *Escherichia coli* B. *Biochemistry* **13**: 300–307.
- Himeno H, Masaki H, Kawai T, Ohta T, Kumagai I, Miura K & Watanabe K (1987) Unusual genetic codes and a novel gene structure for tRNA(AGYSer) in starfish mitochondrial DNA. *Gene* **56**: 219–230.
- Hoels MG & Budisa N (2012) Recent advances in genetic code engineering in *Escherichia coli*. *Curr Opin Biotechnol* **23**: 751–757.
- Huang Y, Wan W, Russell WK, Pai P, Wang Z, Russell DH & Liu W (2010) Genetic incorporation of an aliphatic keto-containing amino acid into proteins for their site-specific modifications. *Bioorg Med Chem Lett* **20**: 878–880.
- Ikeuchi Y, Soma A, Ote T, Kato J, Sekine Y & Suzuki T (2005) Molecular mechanism of lysidine synthesis that determines tRNA identity and codon recognition. *Mol Cell* **19**: 235–246.
- Ikeuchi Y, Kimura S, Numata T, Nakamura D, Yokogawa T, Ogata T, Wada T, Suzuki T & Suzuki T (2010) Agmatine-conjugated cytidine in a tRNA anticodon is essential for AUA decoding in archaea. *Nat Chem Biol* **6**: 277–282.
- Isaacs FJ, Carr PA, Wang HH *et al.* (2011) Precise manipulation of chromosomes *in vivo* enables genome-wide codon replacement. *Science* **333**: 348–353.
- Jaffe JD, Stange-Thomann N, Smith C *et al.* (2004) The complete genome and proteome of *Mycoplasma mobile*. *Genome Res* **14**: 1447–1461.
- Johnson DBF, Xu J, Shen Z, Takimoto JK, Schultz MD, Schmitz RJ, Xiang Z, Ecker JR, Briggs SP & Wang L (2011) RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites. *Nat Chem Biol* **7**: 779–786.
- Jones TE, Brown CL, Geslain R, Alexander RW & de Poupplana Ribas L (2008) An operational RNA code for faithful assignment of AUG triplets to methionine. *Mol Cell* **29**: 401–407.
- Krishnakumar R, Prat L, Aerni H, Ling J, Merryman C, Glass JI, Rinehart J & Söll D (2013) Transfer RNA misidentification scrambles sense codon recoding. *ChemBioChem* **14**: 1967–1972.

- Kwon I, Kirshenbaum K & Tirrell DA (2003) Breaking the degeneracy of the genetic code. *J Am Chem Soc* **125**: 7512–7513.
- Lajoie MJ, Kosuri S, Mosberg JA, Gregg CJ, Zhang D & Church GM (2013a) Probing the limits of genetic recoding in essential genes. *Science* **342**: 361–363.
- Lajoie MJ, Rovner AJ, Goodman DB *et al.* (2013b) Genomically recoded organisms expand biological functions. *Science* **342**: 357–360.
- Lemeignan B, Sonigo P & Marlière P (1993) Phenotypic suppression by incorporation of an alien amino acid. *J Mol Biol* **231**: 161–166.
- Lenski RE, Rose MR, Simpson SC & Tadler SC (1991) Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. *Am Nat* **138**: 1315–1341.
- Lepthien S, Merkel L & Budisa N (2010) *In vivo* double and triple labeling of proteins using synthetic amino acids. *Angew Chem Int Ed Engl* **49**: 5446–5450.
- Li W, Mahapatra A, Longstaff DG, Bechtel J, Zhao G, Kang PT, Chan MK & Krzycki JA (2009) Specificity of pyrrolysyl-tRNA synthetase for pyrrolysine and pyrrolysine analogs. *J Mol Biol* **385**: 1156–1164.
- Link AJ & Tirrell DA (2005) Reassignment of sense codons *in vivo*. *Methods* **36**: 291–298.
- Liu CC & Schultz PG (2010) Adding new chemistries to the genetic code. *Annu Rev Biochem* **79**: 413–444.
- Mandal D, Köhrer C, Su D, Russell SP, Krivos K, Castleberry CM, Blum P, Limbach PA, Söll D & RajBhandary UL (2010) Agmatidine, a modified cytidine in the anticodon of archaeal tRNA(Ile), base pairs with adenosine but not with guanosine. *P Natl Acad Sci USA* **107**: 2872–2877.
- Marlière P, Patrouix J, Döring V, Herdewijn P, Tricot S, Cruveiller S, Bouzon M & Mutzel R (2011) Chemical evolution of a bacterium's genome. *Angew Chem* **123**: 7247–7252.
- Moya A, Gil R, Latorre A, Peretó J, Pilar Garcillán-Barcia M & de la Cruz F (2009) Toward minimal bacterial cells: evolution vs. design. *FEMS Microbiol Rev* **33**: 225–235.
- Mukai T, Kobayashi T, Hino N, Yanagisawa T, Sakamoto K & Yokoyama S (2008) Adding l-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases. *Biochem Biophys Res Commun* **371**: 818–822.
- Mukai T, Hayashi A, Iraha F, Sato A, Ohtake K, Yokoyama S & Sakamoto K (2010) Codon reassignment in the *Escherichia coli* genetic code. *Nucleic Acids Res* **38**: 8188–8195.
- Muramatsu T, Nishikawa K, Nemoto F, Kuchino Y, Nishimura S, Miyazawa T & Yokoyama S (1988a) Codon and amino-acid specificities of a transfer RNA are both converted by a single post-transcriptional modification. *Nature* **336**: 179–181.
- Muramatsu T, Yokoyama S, Horie N, Matsuda A, Ueda T, Yamaizumi Z, Kuchino Y, Nishimura S & Miyazawa T (1988b) A novel lysine-substituted nucleoside in the first position of the anticodon of minor isoleucine tRNA from *Escherichia coli*. *J Biol Chem* **263**: 9261–9267.
- Mursinna RS & Martinis SA (2002) Rational design to block amino acid editing of a tRNA synthetase. *J Am Chem Soc* **124**: 7286–7287.
- Mushegian AR & Koonin EV (1996) A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *P Natl Acad Sci USA* **93**: 10268–10273.
- Nakamura Y, Gojobori T & Ikemura T (2000) Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Res* **28**: 292.
- Nakanishi K, Bonnefond L, Kimura S, Suzuki T, Ishitani R & Nureki O (2009) Structural basis for translational fidelity ensured by transfer RNA lysine synthetase. *Nature* **461**: 1144–1148.
- Neumann H, Peak-Chew SY & Chin JW (2008) Genetically encoding N(epsilon)-acetyllysine in recombinant proteins. *Nat Chem Biol* **4**: 232–234.
- O'Donoghue P, Ling J, Wang Y & Söll D (2013) Upgrading protein synthesis for synthetic biology. *Nat Chem Biol* **9**: 594–598.
- Osawa S, Jukes TH, Watanabe K & Muto A (1992) Recent evidence for evolution of the genetic code. *Microbiol Rev* **56**: 229–264.
- Pezo V, Metzgar D, Hendrickson TL, Waas WF, Hazebrouck S, Döring V, Marlière P, Schimmel P & de Crécy-Lagard V (2004) Artificially ambiguous genetic code confers growth yield advantage. *P Natl Acad Sci USA* **101**: 8593–8597.
- Polycarpo C, Ambrogelly A, Bérubé A, Winbush SM, McCloskey JA, Crain PF, Wood JL & Söll D (2004) An aminoacyl-tRNA synthetase that specifically activates pyrrolysine. *P Natl Acad Sci USA* **101**: 12450–12454.
- Polycarpo CR, Herring S, Bérubé A, Wood JL, Söll D & Ambrogelly A (2006) Pyrrolysine analogues as substrates for pyrrolysyl-tRNA synthetase. *FEBS Lett* **580**: 6695–6700.
- Rennert OM & Anker HS (1963) On the incorporation of 5',5',5'-Trifluoroleucine into proteins of *E. coli*. *Biochemistry* **2**: 471–476.
- Santos MA, Moura G, Massey SE & Tuite MF (2004) Driving change: the evolution of alternative genetic codes. *Trends Genet* **20**: 95–102.
- Schön A, Böck A, Ott G, Sprinzl M & Söll D (1989) The selenocysteine-inserting opal suppressor serine tRNA from *E. coli* is highly unusual in structure and modification. *Nucleic Acids Res* **17**: 7159–7165.
- Schultz DW & Yarus M (1994a) tRNA structure and ribosomal function. I. tRNA nucleotide 27–43 mutations enhance first position wobble. *J Mol Biol* **235**: 1381–1394.
- Schultz DW & Yarus M (1994b) tRNA structure and ribosomal function. II. Interaction between anticodon helix and other tRNA mutations. *J Mol Biol* **235**: 1395–1405.
- Silva FJ, Belda E & Talens SE (2006) Differential annotation of tRNA genes with anticodon CAT in bacterial genomes. *Nucleic Acids Res* **34**: 6015–6022.
- Söll D (1988) Genetic code: enter a new amino acid. *Nature* **331**: 662–663.

- Söll D & RajBhandary U (1995) *tRNA: Structure, Biosynthesis, and Function*. ASM Press, Washington, D.C.
- Söll D & RajBhandary UL (2006) The genetic code – thawing the ‘frozen accident’. *J Biosci* **31**: 459–463.
- Soma A, Ikeuchi Y, Kanemasa S, Kobayashi K, Ogasawara N, Ote T, Kato J, Watanabe K, Sekine Y & Suzuki T (2003) An RNA-modifying enzyme that governs both the codon and amino acid specificities of isoleucine tRNA. *Mol Cell* **12**: 689–698.
- Srinivasan G, James CM & Krzycki JA (2002) Pyrrolysine encoded by UAG in *Archaea*: charging of a UAG-decoding specialized tRNA. *Science* **296**: 1459–1462.
- Suzuki T & Miyauchi K (2010) Discovery and characterization of tRNA<sup>Ile</sup> lysidine synthetase (TilS). *FEBS Lett* **584**: 272–277.
- Suzuki T, Ueda T & Watanabe K (1997) The ‘polysemous’ codon—a codon with multiple amino acid assignment caused by dual specificity of tRNA identity. *EMBO J* **16**: 1122–1134.
- Takemoto C, Spremulli LL, Benkowski LA, Ueda T, Yokogawa T & Watanabe K (2009) Unconventional decoding of the AUA codon as methionine by mitochondrial tRNA<sup>Met</sup> with the anticodon f<sup>5</sup>CAU as revealed with a mitochondrial *in vitro* translation system. *Nucleic Acids Res* **37**: 1616–1627.
- Tang Y & Tirrell DA (2002) Attenuation of the editing activity of the *Escherichia coli* leucyl-tRNA synthetase allows incorporation of novel amino acids into proteins *in vivo*. *Biochemistry* **41**: 10635–10645.
- Taniguchi T, Miyauchi K, Nakane D, Miyata M, Muto A, Nishimura S & Suzuki T (2013) Decoding system for the AUA codon by tRNA<sup>Ile</sup> with the UAU anticodon in *Mycoplasma mobile*. *Nucleic Acids Res* **41**: 2621–2631.
- Tuite MF & Santos MA (1996) Codon reassignment in *Candida* species: an evolutionary conundrum. *Biochimie* **78**: 993–999.
- Wang L & Schultz PG (2004) Expanding the genetic code. *Angew Chem Int Ed Engl* **44**: 34–66.
- Wang L, Brock A, Herberich B & Schultz PG (2001) Expanding the genetic code of *Escherichia coli*. *Science* **292**: 498–500.
- Wang P, Fichera A, Kumar K & Tirrell DA (2004) Alternative translations of a single RNA message: an identity switch of (2S,3R)-4,4,4-trifluorovaline between valine and isoleucine codons. *Angew Chem Int Ed Engl* **43**: 3664–3666.
- Wang L, Xie J & Schultz PG (2006) Expanding the genetic code. *Annu Rev Biophys Biomol Struct* **35**: 225–249.
- Wolstenholme DR (1992) Animal mitochondrial DNA: structure and evolution. *Int Rev Cytol* **141**: 173–216.
- Young TS, Ahmad I, Yin JA & Schultz PG (2010) An enhanced system for unnatural amino acid mutagenesis in *E. coli*. *J Mol Biol* **395**: 361–374.
- Zhang G, Ignatova Z & Idnurm A (2009) Generic algorithm to predict the speed of translational elongation: implications for protein biogenesis. *PLoS One* **4**: e5036.

## Supporting Information

Additional Supporting Information may be found in the online version of this article:

**Data S1.** Experimental details.

**Fig. S1.** Schematic presentation of the pNB26'2 cargo site.

**Fig. S2.** Comparison of growth curves between *Escherichia coli* MG1655 WT (black line) and *E. coli* MG1655  $\Delta$ tilS: cat (pNB26'2\_MmPair) (red line) including standard deviation error bars.

**Fig. S3.** Semi-logarithmic plot of OD<sub>600 nm</sub> values over time.

**Table S1.** Sequences of confirmation primers used in verification colony PCR.